DOCUMENT RESUME

ED 349 868 HE 025 785

AUTHOR Smith, Constance K.; Waggener, Anna T.
TITLE Utilizing Throw-Away Data: Invalid and Missing Data Can Have Meaning! AIR 1992 Annual Forum Paper.
PUB DATE May 92
NOTE 15p.; Paper presented at the Annual Forum of the Association for Institutional Research (32nd, Atlanta, GA, May 10-13, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Academic Persistence; College Applicants; College Freshmen; Data Collection; Enrollment; *Error Patterns; Higher Education; *Predictor Variables; *Questionnaires; Research Design; *Response Rates (Questionnaires); Surveys
IDENTIFIERS *AIR Forum; Southeastern Louisiana University

ABSTRACT

Two studies were conducted at the Southeastern Louisiana University (SLU) to determine possible uses of errors and omissions on surveys of incoming and freshmen college students. The subjects of the first study were 1,927 individuals who had applied for admission to SLU and attended freshman orientation in the summer of 1989. Blanks and incorrect responses to a survey were taken as an independent variable, GOOF. A t-test found that the mean GOOF score of the respondents who enrolled that semester was significantly lower than the mean GOOF score of respondents who did not enroll. The second study involved a random sample of 1,540 new freshmen attending SLU in the fall of 1989. Two independent variables were the GOOF variable and one called RESPDNT for those subjects who did not return a mailed survey. Dependent variables were subsequent academic performance and fall-to-fall retention. Results indicated that for enrolled students the GOOF variable may not be an important one, but failure to respond (RESPDNT) to a survey may have meaning in terms of academic performance and retention. Included are appended instructions for creating the GOOF variable and six references. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

#23

ED349868

Utilizing Throw-Away Data:

Invalid and Missing Data Can Have Meaning!

Constance K. Smith, Ph.D.

Anna T. Waggener, Ph.D.

Southeastern Louisiana University
P.O. Box 490, SLU
Hammond, LA 70402
504-549-2077

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE 025 785

Paper Presented to
The Association of Institutional Research
Atlanta, GA
May 10-13, 1992

2 BEST COPY AVAILABLE

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Second Annual Forum of the Association for Institutional Research held at the Atlanta Hilton & Towers, Atlanta, Georgia, May 10-13, 1992. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Jean Endo
Chair and Editor
Forum Publications
Editorial Advisory Committee

Utilizing Throw-Away Data:
Invalid & Missing Data **CAN** Have Meaning!

Abstract

When they encounter blanks and invalid responses in survey data, researchers routinely code them as "missing values". As a result, such data usually are excluded from the data analysis. Similarly, failure to respond to a survey frequently is assumed to have no meaning.

It seems possible, however, that when a potential college student or a new freshman fails to respond or provides incomplete or invalid survey data, these behaviors might have some meaning in terms of subsequent enrollment, academic performance and retention.

A simple process of coding can change invalid responses and blanks into a new variable (called "GOOF"). The relationship of this variable *vis a vis* retention is tested in two different studies. One of these studies also explores the relationship of failure to respond to a freshman survey to the student's subsequent retention and academic performance.

Utilizing Throw-Away Data: Invalid and Missing Data Can Have Meaning!

Psychologist Jean Piaget developed his theory of cognitive development by noticing the errors which children systematically make at various ages (Piaget, 1966). Unlike most developmental theorists, he took errors seriously -- as an indication of the developing child's cognitive structure.

Retention studies frequently cite preparation and motivation or commitment as powerful factors in determining whether or not students persist in college (Douzenis, 1990; Kinnick & Kemper, 1988; Tinto, 1987). Preparation usually is determined by high school performance and/or scores on standardized college admissions tests. Motivation or commitment might also be measured by items on a survey instrument (Stage, 1988). Tinto (1987) emphasized that the initial intent of the student regarding his or her educational participation is a strong predictor of persistence or attrition. Additionally, Bean (1982) stated that if institutions survey educational goals and commitments of their incoming students, then the institution can more accurately predict persistence or withdrawal.

Surveys of incoming students, then, might provide crucial information relevant to persistence. Survey researchers know, however, that some students fail to respond to surveys at all and other students respond, but leave blanks and make errors on their forms. Generally, non-respondents, blanks and invalid data do not enter into the data analysis

process. Blanks and invalid responses are coded as "missing values" and then ignored. We literally throw them away!

Nevertheless, a student's failure to respond or providing incomplete and/or invalid responses might tell us something about that student's motivation level and/or commitment to higher education. Perhaps by exploring errors and omissions like Piaget did, survey researchers can glean more information from their data than they had anticipated. These theories prompted two studies to determine if non-response and/or frequent errors and blanks are related to subsequent enrollment, to academic performance and/or to retention. Both of these studies were conducted on the campus of Southeastern Louisiana University in Hammond, Louisiana.

Study One

Methodology

The subjects in the first study were 1,927 individuals who had applied for admission to Southeastern Louisiana University and who attended freshman orientation in June, July, or August of 1989.

The Supplementary Enrollment Information instrument (SEI) used to collect the data was designed to measure these potential students' characteristics, goals, and attitudes toward self, family, and educational commitment. It was integrated into the "final exam" which was administered at the end of freshman orientation. Some of the orientation

participants might not have completed the SEI instrument, but we have no way of identifying them.

The GOOF variable used in the study was created by recoding blanks and incorrect responses (see Appendix A). This independent variable (GOOF) was subsequently analyzed with the dependent variables of enrollment and fall-to-fall retention.

The first purpose of this study was to determine if the frequency of invalid and/or incomplete responses to the SEI instrument (GOOF variable) was related to whether or not the respondents enrolled at Southeastern that Fall semester. The second purpose was to determine if the GOOF score on SEI was related to whether or not respondents re-enrolled their second Fall semester.

Independent t-tests were used to test the differences between groups[1]. The $p < .05$ level of significance was used in all tests.

Results

In testing the difference between respondents who enrolled in Fall, 1989 and respondents who did not enroll in the same semester regarding the GOOF variable, an Independent t-test was used to compare the two different groups. This t-test found that the mean GOOF score of the respondents who enrolled that semester was significantly ~~higher~~ lower than the mean GOOF score of the respondents who did not enroll ($t_{(2,260)} = 2.58$, $p < .05$).

---

[1] The Statistical Package for the Social Sciences (SPSS) was the software used. The mainframe VAX computer at Southeastern Louisiana University (Hammond) was used for the statistical computation.

Groups	Mean GOOF
Did Enroll in University	0.2808
Did NOT Enroll	0.8489

Another independent t-test was used to analyze the difference between students who re-enrolled the second Fall semester after the freshman orientation and students who did not re-enroll the second Fall semester after the same orientation. This t-test found that the mean of the students who returned a second semester was not significantly different than the mean of the students who did not return for a second Fall semester ($t_{(2,035)}$= .33, n.s.).

Discussion

The Statistical Package for the Social Sciences (SPSS) ignores unlabeled missing values. When these missing values were recognized by the researcher and then recoded, these data were meaningful in terms of subsequent enrollment at the University.

Students who had lower GOOF scores on the SEI instrument were more likely to enroll the subsequent Fall semester than were students who had higher GOOF scores. Low GOOF scores might suggest that those students showed more attention to detail, more interest in college preparation, and a deeper commitment to a college education than those students who showed more blanks. However, the frequency of GOOFs on the SEI instrument did not impact the likelihood that a student will return the following Fall semester.

Study Two

Methodology

The subjects in the second study were a random sample (n = 1,540) of the new freshman attending Southeastern during the fall semester of 1989. The instrument used to collect information was designed to measure these students' expectations of the university, educational and life goals, and other aspects of their experiences as new students. The survey was administered via mail, and over 63% of the sample responded.

In addition to the information supplied by the respondents, two data elements were created. First, whether or not the student responded to the survey was made into a dichotomous variable called RESPDNT. Second, for those 967 freshmen who did respond to the survey, blanks were converted into a data element called GOOF.

Both of these new data elements (RESPDNT and GOOF) were the independent variables in this second study. The dependent variables were subsequent academic performance and fall-to-fall retention.

The purposes of this second study were to determine if (1) if there was a significant relationship between respondents' GOOF scores and their cumulative grade point averages; (2) if the GOOF scores of students who were retained were significantly different from the GOOF scores of students who were not retained; (3) if the cumulative grade point averages of freshmen who responded to the survey were significantly different from the CumGPA's

of freshmen who did not respond to the survey; and (4) if the retention rate of respondents was significantly different from the retention rate of non-respondents.

The statistical tests used were: Pearson correlation, one-way analysis of variance and chi square[2]. The $p < .05$ level of significance was used in all tests.

Results

GOOF (e.g. the total number of errors/blanks per respondent) was not related to cumulative grade point average at the end of the freshman year (Pearson $r = -.0127$, n.s.). It also was not related to retention either the following Spring semester ($F_{(1,965)} = 0.5589$, n.s.) or the Fall semester of the second year ($F_{(1,965)} = 1.3926$, n.s.).

However, freshmen who responded to the survey subsequently had significantly higher cumulative grade point averages than did freshmen who failed to respond ($F_{(1,948)} = 38.85$, $p < .0000$).

Respondent Groups	Mean CumGPA
Responded to Fr. Survey	2.4691
Did NOT Respond to Survey	2.1668

In addition, 1989 freshmen who responded to the survey were more likely than non-respondents to be retained the Fall semester one year later ($X^2_{(1,1560)} = 31.21$, $p < .0000$).

[2] The Statistical Package for the Social Sciences (SPSS) was the software used. The mainframe VAX computer at Southeastern Louisiana University (Hammond) was used for the statistical computation.

Respondent Groups	Not Retained F'90	Retained F'90
Responded to Fr. Survey	32.4%	67.6%
Did NOT Respond to Survey	46.5%	53.5%

Discussion

For enrolled students, GOOF might not be an important variable. Failure to respond to a survey, however, might have meaning in terms of academic performance and retention. It should be noted, however, that our survey procedures included persistent follow-ups of non-respondents. Students who failed to respond not only ignored the initial survey mailing, but three reminders and a second survey mailing as well.

Summary and Implications

These two studies were based on the hunch that failure to respond to a survey or providing incomplete or invalid data (GOOF) might mean something. The most likely meanings might be cognitive, motivational *vis a vis* the immediate task, or commitment to the University.

Prior to enrollment, the GOOF variable seems to differentiate between applicants who will register from those who will not. Since GOOF seems not to be related to subsequent academic performance or to retention, it might not reflect either the respondent's cognitive abilities or (once he/she has enrolled) his/her commitment to

continuing at the institution. Rather, GOOF might reflect an applicant's commitment to completing the enrollment process.

Once the applicant is enrolled as a freshman, however, invalid or incomplete responses seem to have no meaning. On the other hand, willingness to respond to a survey[3] does seem to have meaning in terms of subsequent academic performance and retention. Non-compliance in completing a survey could be related to a general attitude of non-compliance with regard to academic demands or it could reflect a lack of commitment to continued attendance at the institution.

Researchers at open-admissions institutions might want to take advantage of the additional information which GOOF and failure-to-respond can provide in studying success and retention at various points in the enrollment process and during the initial two years of college.

[3] Assuming a reasonably assertive follow-up procedure is used to maximize response rate

References

Bean, J. P. (1982). Conceptual models of student attrition: How theory can help the institutional researcher. In E. T. Pascarella (Ed.), Studying Student Attrition, 17-33, New Directions for Institutional Research #36, San Francisco: Jossey-Bass, Inc.

Douzenis, C. (1990). Characterizing student attrition: A comparison of transfers, dropouts and continuing students. Unpublished paper, Association for Institutional Research Forum.

Kinnick, M. K. and Kemper, K. (1988). Beyond "front door" access: Attaining the bachelor's degree. Research in Higher Education, 29 (4), 299-317.

Piaget, Jean (1966). Origins of intelligence in children, (Cook, Margaret, tr.). Madison, CT: International Universities Press, Inc.

Stage, F. K. (1988). University attrition: LISREL with logistic regression for the persistence criterion. Research in Higher Education, 29 (4), 343-357.

Tinto, V. (1987). Leaving college: Rethinking the causes and cures of student attrition. Chicago: University of Chicago Press.

HOW TO CREATE A "GOOF" VARIABLE

For numeric variables, SPSS defaults blank values into "missing".

To count a blank as a GOOF, you must pretend that you have string variables rather than numeric variables.

In your DATA LIST, insert **(A)** after each numeric variable you want to turn into a string variable.

For example, the DATA LIST for a four-question survey might be:

/Q1 14-15 Q2 16-17 Q3 18 Q4 19

To change to string variables, revise it to:

/Q1 14-15 **(A)** Q2 16-17 **(A)** Q3 18 **(A)** Q4 19 **(A)**

Now you can RECODE the string variables so that blanks and invalid responses can be counted.

Suppose, on the first two questions of your survey, valid responses could range from 1 - 10. A blank is invalid, a zero is invalid and anything over 10 is invlaid.

In order to count the blanks and invalid values, the recode command would read:

```
RECODE Q1 to Q2 ('ß1','ß2','ß3','ß4','ß5','ß6','ß7','ß8',
                'ß9','10'=0)(else=1) into G1 to G2.
```

Suppose on the next two questions of your survey, valid responses could range from 1 - 4. The recode command would read:

```
RECODE Q3 to Q4 ('1','2','3','4'=0)(else=1) into
                                    G3 to G4.
```

Then all you need to do is add up your new variables:

```
COMPUTE GOOF=G1+G2+G3+G4.
```

GOOF is a numeric variable which you can analyze as you would AGE or CUMULATIVE GPA.

[Note: ß means leave a space.]